

Augmented reality visualization of scene depth for aiding ROV pilots in underwater manipulation

Fabio Bruno, Antonio Lagudi, Loris Barbieri*, Domenico Rizzo, Maurizio Muzzupappa, Luigi De Napoli

University of Calabria, Department of Mechanical, Energy and Management Engineering (DIMEG), Via Pietro Bucci, 87036, Arcavacata di Rende, CS, Italy

ARTICLE INFO

Keywords:

Underwater manipulation
Forward kinematics
Optical-stereo camera
Augmented reality

ABSTRACT

Underwater manipulation is a key technology for marine industries and exploration that can be efficiently adopted in other application fields, such as underwater archaeology, biological manipulation, scientific expedition, as well as offshore construction in the Oil and Gas industry. It is performed remotely by expert pilots thanks to the visual feedbacks provided by one or more cameras but without any information about the distance between the end-effector and the target.

To this end, the paper presents a novel system based on a sensorized robotic arm, stereoscopic 3D perception and augmented reality visualization to support ROV's pilots in underwater manipulation tasks. The system, thanks to the adoption of an optical-stereo camera, provides a visual feedback of the underwater scene on which a depth map of the underwater workspace is augmented on. In particular, combining the kinematics of the robotic arm and the standard photogrammetric model of the stereo camera, it is possible to generate a depth map that shows to the pilots the distances of the surface of the scene objects from the end-effector's pose. Experimental tests carried out in the context of the CoMAS (In-situ conservation planning of Underwater Archaeological Artefacts) project have demonstrated the effectiveness of the proposed system.

1. Introduction

Underwater manipulation conducted in shallow and deep water is an essential operation for performing underwater works in several application fields like offshore construction and ocean engineering. These operations are usually performed by means Remotely Operated Vehicles (ROVs) that are remotely controlled thanks to the visual feedbacks provided by one or more cameras that allow pilots to estimate the morphology of the submerged environment (Zhang et al., 2017; Shim et al., 2010). Since the quality of these feedbacks is strongly affected by many factors, underwater manipulation becomes a very complex and tricky operation that requires considerable experience of the pilot (Sagara and Ambar, 2015; Sakagami et al., 2010). To this end, the attention of the scientific community in the field of Underwater Vehicle Manipulator Systems (UVMS) is focusing on the development of methods and techniques to improve the effectiveness of underwater manipulation (Antonelli et al., 2004; Mohan and Kim, 2015; Galloway et al., 2016). In fact, advanced sensors, such as sonars and camera motion tracker systems (Marani and Yuh, 2009) or structured light

techniques (Prats et al., 2012) have been proposed in the last years. Although the substantial contribution to scientific innovation from these studies, experiments in the field remain dependent on the operator capacity to pilot UVMS. This limit appears more evident in some emerging application fields, such as the underwater cultural heritage (UCH), and specifically for the cleaning and restoration of submerged structures and artefacts, where high dexterity and soft touch of the end-effector arm are necessary requirements. If, on the one hand, underwater manipulation represents a great opportunity for UCH field, on the other hand, it needs specific adjustments and improvements in order to meet marine archaeology's requirements. In fact, underwater manipulators are mainly designed for undersea construction and pipeline installation and maintenance with the consequence that the adoption of these robotic arms can end up damaging or destroying valuable and fragile artefacts.

To this end, the paper proposes an augmented reality (AR) visualization system which represents a new technological solution for the UCH field to efficiently support and facilitate the manipulation of archaeological artefacts and, more specifically, the planned maintenance

* Corresponding author.

E-mail addresses: fabio.bruno@unical.it (F. Bruno), antonio.lagudi@unical.it (A. Lagudi), loris.barbieri@unical.it (L. Barbieri), domenico.rizzo@unical.it (D. Rizzo), muzzupappa@unical.it (M. Muzzupappa), luigi.denapoli@unical.it (L. De Napoli).

required to prevent the biological colonization of the submerged structures. Furthermore, the proposed system can be efficiently applied to marine science research activities, especially when collecting fragile biological samples.

In particular, the paper presents a novel system based on a sensorized robotic arm, stereoscopic 3D perception and augmented reality visualization to support ROV's pilots in underwater manipulation tasks. The system provides a visual feedback of the submerged scene on which a depth map of the underwater workspace is augmented on thanks to the adoption of an optical-stereo camera. Combining the kinematics of the robotic arm and the standard photogrammetric model of the camera (Brown, 1971) it is possible to provide to the pilots the depth map with respect to the end-effector's pose.

This work has been carried out in the context of the CoMAS (In-situ conservation planning of Underwater Archaeological Artefacts) project ([CoMAS project, 2013](#)) that aimed to discover and develop new materials, techniques and methodologies for the conservation and restoration of marine sites in their natural environment ([Bruno et al., 2016, 2015a; 2015b](#)).

The paper is organized as follows. Section 2 provides an overall description of the system architecture. Section 3 focuses on the mathematical model adopted for the depth map calculation. In particular, the depth map is computed with respect to the end-effector's pose by taking into account the kinematic model of the robotic arm (Section 4) and the standard photogrammetric model of the optical-stereo camera (Section 5). Finally, laboratory and field tests are described in Section 6.

2. System architecture

The system architecture consists of a ROV ([Fig. 1](#)) equipped with a custom-made sensorized underwater robotic arm and an optical-stereo camera for a stereoscopic 3D perception of the underwater scene, and a software specifically developed to provide to ROV's pilots an augmented reality visualization of the depth map of the underwater framed scene calculated with respect to the end-effector's pose.

The system has been designed for the CoMAS project to provide different degrees of automation for the various mechanical cleaning

activities to be carried out in an archaeological site, and therefore to simplify the workflow by means of semi-automatized activities.

2.1. ROV

The ROV is a Perseo GTV by Ageotec company. It is a midsize multi-purpose vehicle optimized to perform both the visual and instrumental inspection and the manipulation of small- or medium-sized objects. The underwater vehicle is built around a supporting chassis in polypropylene with stainless steel brace for thrusters stiffening, lifting and fixing. The underwater vehicle's propulsion derives from six thrusters driven by brushless electric motors, and it features auto-heading and auto-depth functions to move while keeping the proper path at a constant depth. The vehicle has been equipped with the following sensors: IMU Xsens MTI 300, an AA EasyTrack USBL, and a Link Quest 600 Micro DVL.

2.2. Underwater robotic arm

The underwater arm ([Fig. 2](#)) has been re-engineered starting from a commercial 5-functions hydraulic arm, the Hydro-Lek HLK-43000, which incorporates a continuous jaw rotate mechanism with an integrated cutter tool. The manipulator has been modified to ensure a greater working volume to easily perform mechanical cleaning activity in the underwater archaeological context. The original gripper tool has been removed and the continuous jaw rotating mechanism has been modified in a pan-tilt system to handle an electromechanical brush tool ([Barbieri et al., 2017](#)).

The arm has been sensorized with three Acurro AR-63 waterproof CAN-bus magnetic encoders, for measuring the angular position of each link, and a couple of RE22 SSI magnetic encoders, for the measuring of the pan-tilt head movements. While the electromechanical brush has been equipped with three sensors S2Tech BC302-20 kg for measuring the axial forces. The end-effector uses a 1200W brushed motor, with a maximum torque of 15Nm at 1250 rpm, and can handle different types of brushes to perform all the planned maintenance operations. A human operator uses two joysticks to control the robotic arm using direct



Fig. 1. The ROV equipped with the custom-made robotic arm and optical-stereo camera.

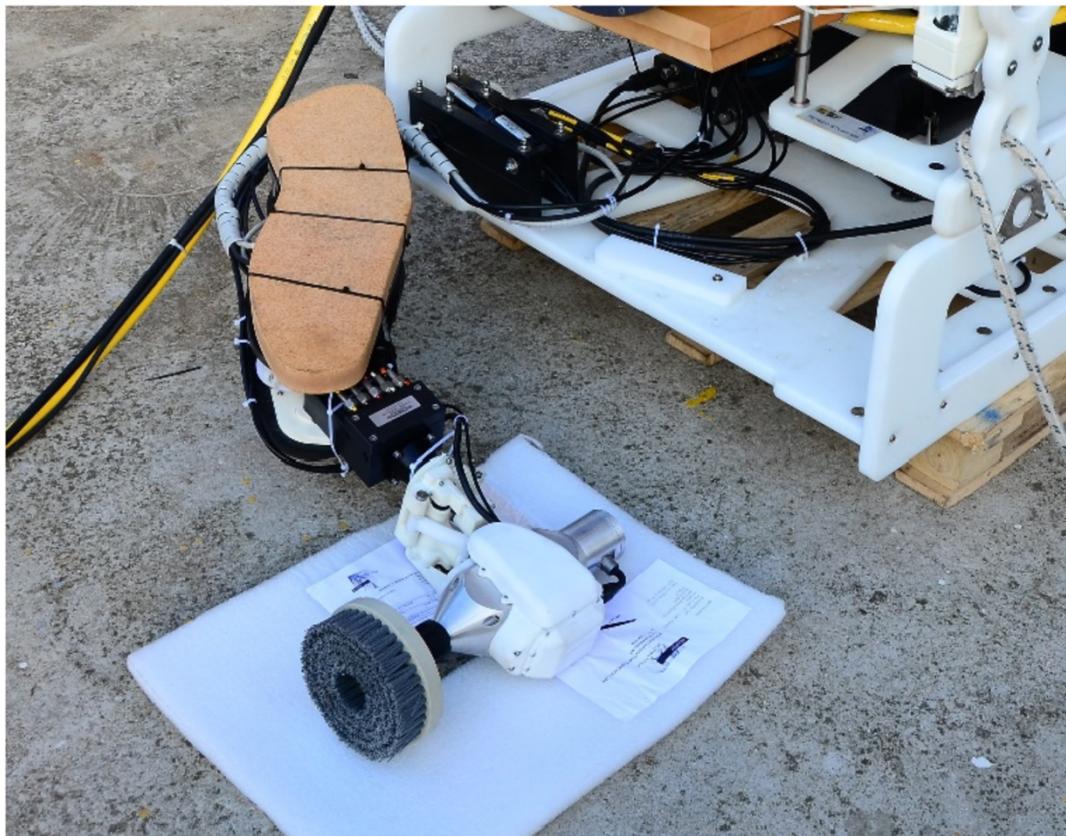


Fig. 2. The re-engineered underwater arm mounted on the skid of the ROV.

kinematic method. Nevertheless, since the robotic arm has been sensorized with magnetic encoders that provide feedback about the position of each link of the arm also a master-slave control could be implemented.

2.3. Optical-stereo camera

The optical-stereo camera (Fig. 3) is the result of a research activity conducted in the field of the underwater stereo photogrammetry, both for passive and active applications (Bruno et al., 2011; Bianco et al., 2013; Skarlatos et al., 2010).

It consists of two cameras, in a stereo configuration, housed in custom-made waterproof cases. The digital cameras are two ultra-compact Imaging Source DFK 23G445 GigE color industrial camera with a 1/3-inch Sony CCD sensor format, a resolution of 1280×960 pixels, and a frame rate of 30 fps. They are equipped with a pair of Theia Technologies SL183A lenses that provide an ultra-wide field of view, with a distortion less than 1%, thanks to a focal length ranged from 1.8 to 3.0 mm and aperture F1.8. The custom-made waterproof cases are made of aluminum to ensure an efficient heat dissipation, while the flat port of the camera housings is made of polycarbonate.

The optical-stereo camera is mounted on the skid of the ROV (Fig. 1) in order to be adopted also for the navigation with the arm retracted in its home position. The adoption of the current configuration is due to the fact that the optical-stereo camera is too bulky and heavy to be mounted on the robotic arm. Furthermore, if placed on the manipulator, the camera becomes particularly vulnerable insofar as it is more subject to eventual impacts and collisions with the underwater scenario.

2.4. Augmented reality visualization

The underwater scene captured by the optical-stereo camera is

furnished to the ROV operators by means of a software developed ad-hoc. The software processes the raw images acquired by the stereoscopic camera to create layers of 3D information that are overlaid, by means of augmented reality techniques, to the main 2D video feedback. In particular, Libelas and OpenCV (OpenCV, 2014) libraries have been adopted to perform an on-line augmented reality visualization of the 3D underwater scene. OpenCV library has been used for the rectification of the stereo pair, while Libelas library has been implemented in the software both for the stereo matching operation and the generation of the disparity map. The adoption of Libelas library allows to generate 3D point clouds, of about 200,000 points, at a frame rate of 7 fps. Furthermore, its Efficient LArge-scale Stereo (ELAS) algorithm (Geiger et al., 2011) allows for obtaining a more robust and accurate 3D point cloud in the underwater environment.

The software devoted to the AR visualization of the scene depth perception has been developed to feature a high customizability of user interfaces and a high flexibility of parameters to allow users to set the main parameters of the optical-stereo camera dependently by the conditions of the underwater environment. The following picture (Fig. 4) depicts the depth map, of the distance of the end-effector from a submerged archaeological remain, augmented on the images of the underwater scenario acquired by the optical-stereo camera. In particular, the 3D point cloud of the underwater scene is shown on the right side of the screen, whereas, a stereo image pair of the acquired frame and the 3D depth map of the scene are broadcasted respectively on the upper and bottom left side of the screen.

Two methods have been embedded in the software for the calibration of the optical-stereo camera. The first one is the well-known calibration method proposed by Zhang (1999), the second one is the calibration method developed in (Rahman and Krouglicof, 2012) that offers robust performances regardless of the degree of lens distortion and therefore has better performance in underwater applications.

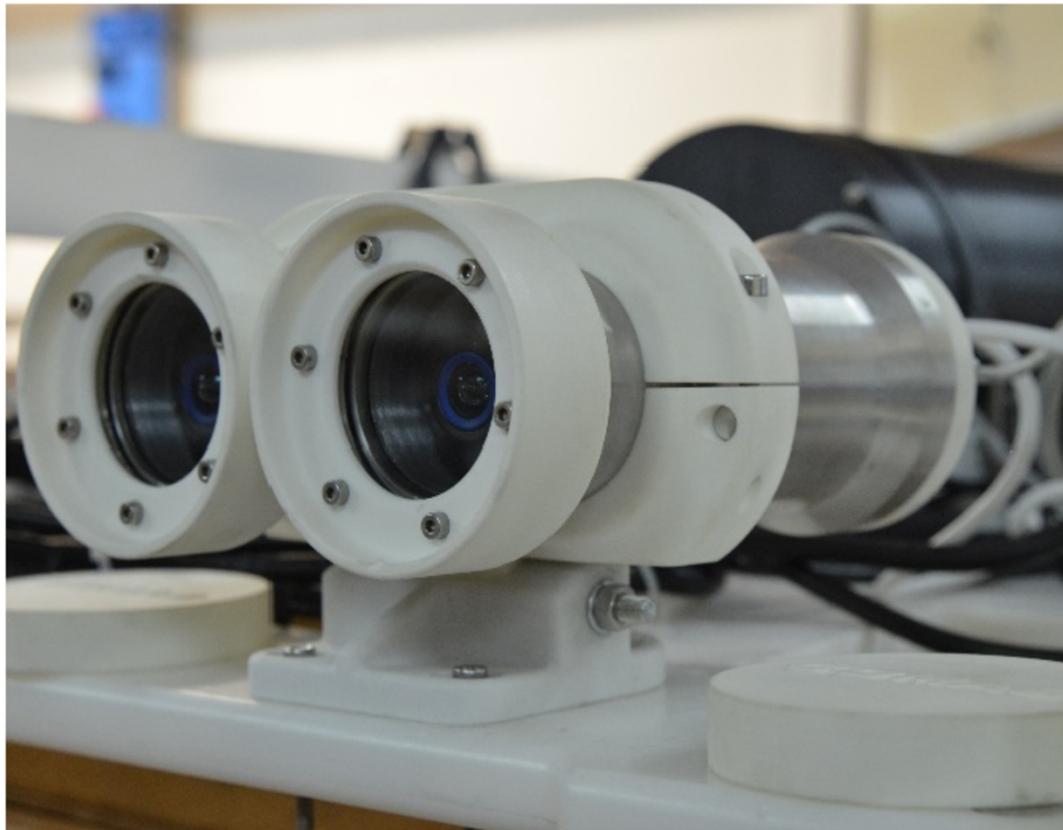


Fig. 3. Optical-stereo camera.

3. Mathematical model of the system

The following figure (Fig. 5) depicts the mathematical model adopted to generate a depth map in which the distance of a generic point P, acquired by the optical-stereo camera, is calculated with respect to the end-effector's position. The mathematical formulation considers both the direct kinematics of the robotic arm and the standard photogrammetric model of the optical-stereo camera, consisting of a pinhole camera with additional parameters (radial, decentring distortions, affinity and shear). In particular, {E}, {C} and {O} are the local

reference frames respectively of: the end-effector's tool, the optical-stereo camera, and the robotic arm.

Based on the above-mentioned mathematical model, the position of a generic point P can be calculated with respect to the end-effector's reference frame {E} by means of the following formulation:

$${}^E P = {}^E T_0 {}^0 T_C {}^C P \quad (1)$$

where ${}^E T_0$ is calculated by means of the forward kinematic model of the robotic arm and it represents the end-effector pose with respect to the frame {O}; ${}^0 T_C$ is the homogeneous transformation matrix of the frame

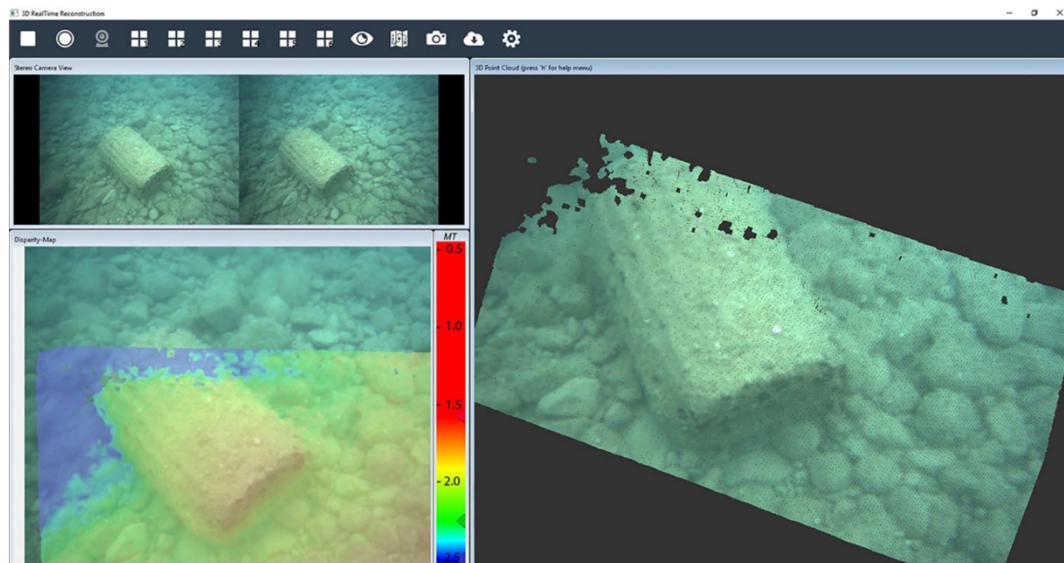


Fig. 4. Control software for the AR scene depth visualization.

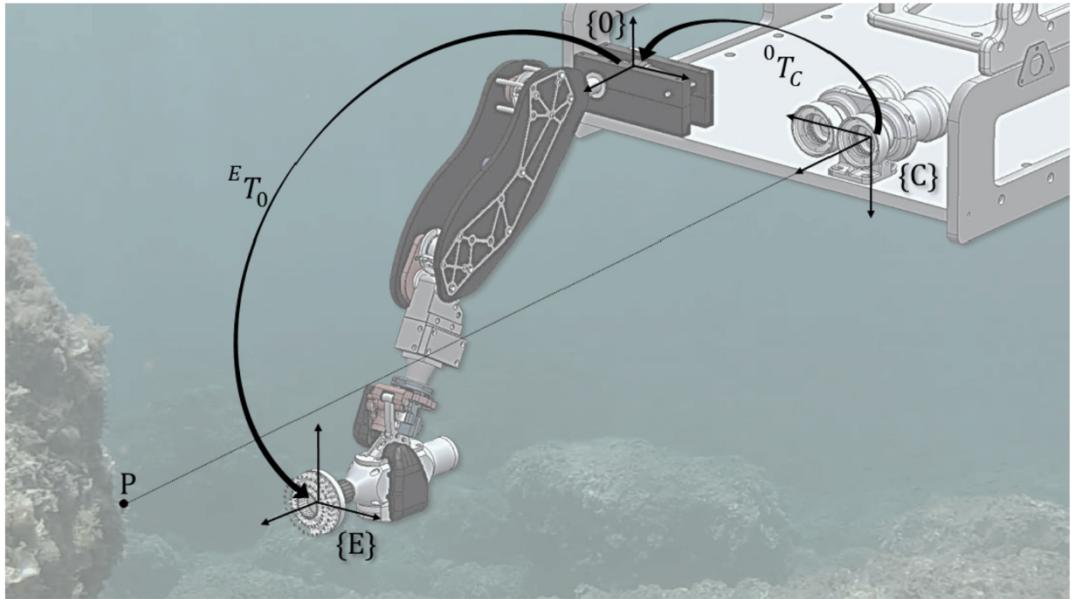


Fig. 5. Mathematical model representation of the proposed system.

{0} with respect to the frame {C}; and c_P is the position vector of the generic point P referred to the coordinate system {C}.

4. Kinematic model of the arm

The mathematical description of the kinematic chains of the 5-DOF underwater manipulator has been defined according to the DH-parameter notation (Denavit and Hartenberg, 1955).

The DH notation allows to identify the DOFs and number the links starting from the basement, referred as link 0, of the robotic arm. As depicted in Fig. 6, the robotic arm has five links and five joints, four of which are revolute joints, with rotation along the z-axes, and the last one, represented within the orange box, is the tilt mechanism. The following table (Table 1) summarizes the DH-parameters of the forward kinematics of the first four joints established on the basis of the zero-reference configuration (Fig. 6). In particular: $i^{-1}T_i$ are the homogeneous transformation matrices which describe the position of each joint frame (i) with respect to the preceding frame (i-1); θ_i are the z-axis rotations; d_i are the frame's offsets along the z-axis; a_i are the link

Table 1
DH-parameters.

$i^{-1}T_i$	θ_i [deg]	d_i [mm]	a_i [mm]	α_i [deg]	Limit q_i [deg]
0T_1	q_1	229	0	90	20 to -90
1T_2	q_2	0	450	180	-10 to 80
2T_3	$q_3 + 90$	0	0	-90	0 to 45
3T_4	q_4	225	0	0	0 to 180

lengths along the x-axis; and α_i are the x-axis rotations.

Regarding the resolution of the direct kinematics of the tilt mechanism (Fig. 7), it is a quadrilateral mechanism that can be expressed as follows:

$${}^4T_5 = Dz(b) \cdot Ry(\delta) \cdot Dz(c/2) \cdot Ry(-\delta) \cdot Ry(-\phi) \cdot Ry(q_5) \quad (2)$$

where b, δ , c and ϕ are the parameters of the four-bar linkage, and the q_5 angle rotation can vary between 0 and 90°.

Similarly, the last transformation matrix referred to the end-effector

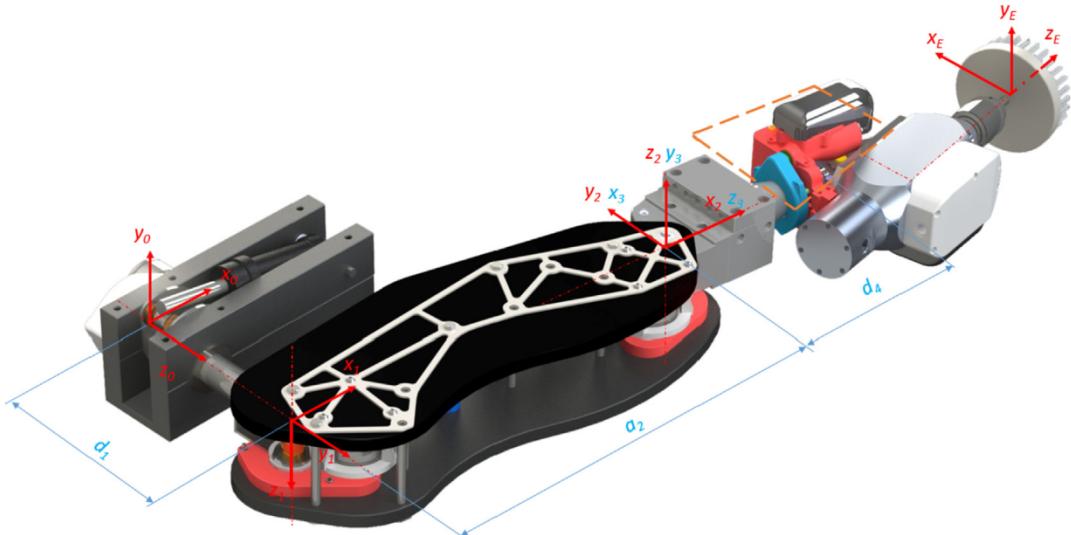


Fig. 6. Zero-reference configuration of the robotic arm and DH notation.

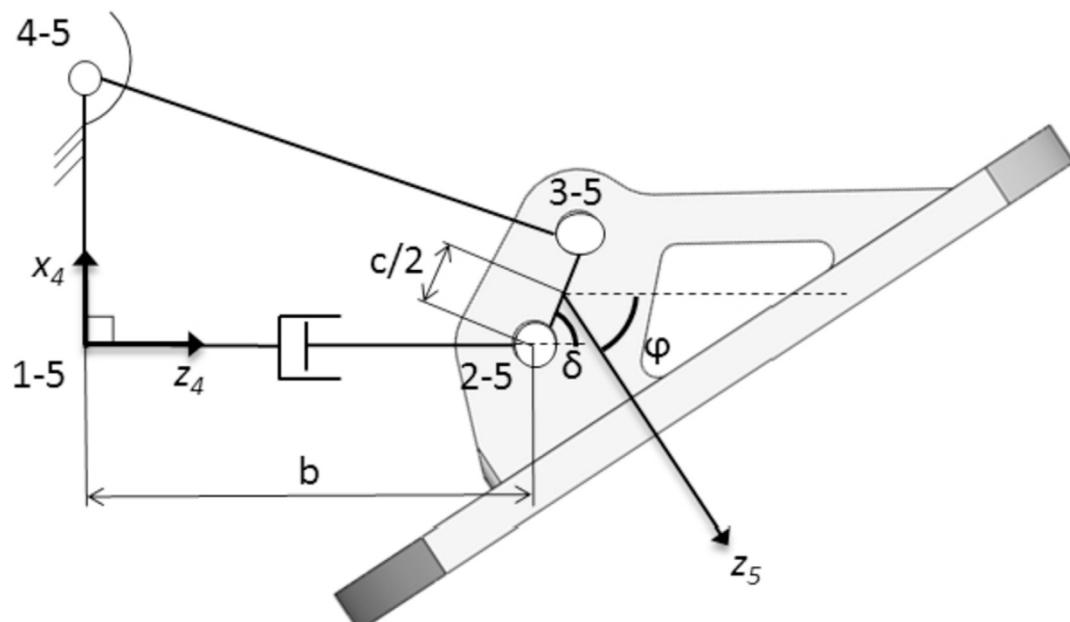


Fig. 7. Schematic representation of the quadrilateral mechanism and characteristic parameters.

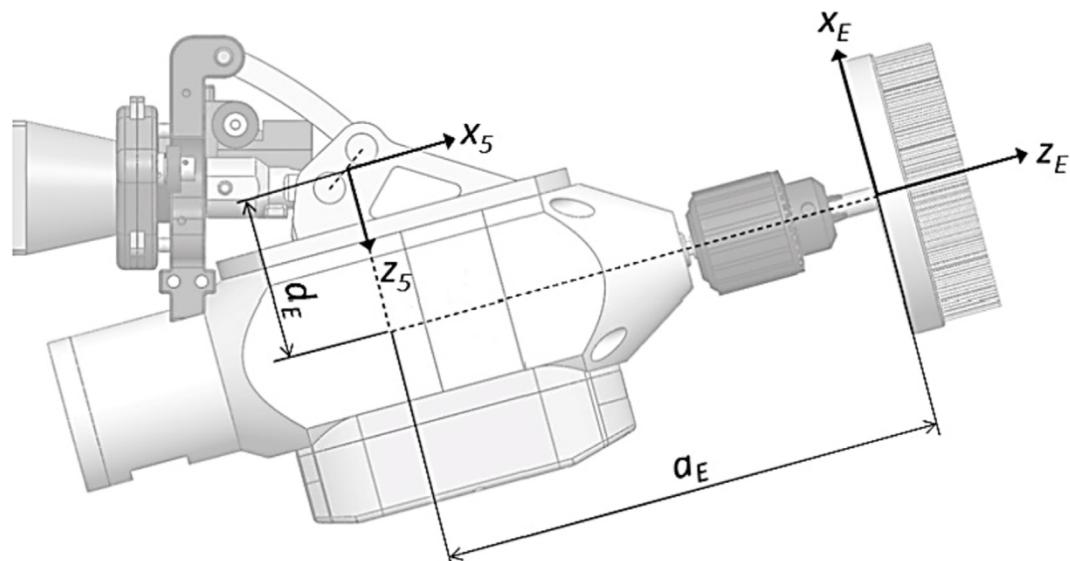


Fig. 8. Local reference frame and geometric parameters of the end-effector tool.

of the robotic arm (Fig. 8) has been formulated as follows:

$${}^5T_E = Dz(d_E) \cdot Dx(a_E) \cdot Ry(90^\circ) \quad (3)$$

where \$d_E\$ is the frame's offset along the \$z\$-axis and \$a_E\$ is the link length along the \$x\$-axis.

On the basis of the above-mentioned considerations, the kinematic model of the underwater manipulator, that specifies the relative pose of the end-effector and the base coordinate system, is expressed by means of the homogeneous transformation matrix:

$${}^0T_E = \begin{bmatrix} {}^0R_E & S \\ 0_{3 \times 3} & 1 \end{bmatrix} = {}^0T_1 \cdot {}^1T_2 \cdot {}^2T_3 \cdot {}^3T_4 \cdot {}^4T_5 \cdot {}^5T_E \quad (4)$$

where the submatrix \${}^0R_E\$ and the array \$S = [x \ y \ z]^T\$ are respectively rotational matrix and position vector of the end-effector.

5. Camera to arm pose estimation

The estimation of the pose between the optical-stereo camera and the robotic arm, i.e., \${}^0T_C\$ transformation matrix, has been experimentally determined by means of the geometric approach illustrated in the following figure (Fig. 9). Since the camera reference frame {C} is a theoretical point not known a priori, then a new reference frame {C'} has been added into an easily identifiable and accessible point located on the external waterproof case of the camera. As a consequence, the unknown rigid transformation matrix \${}^0T_0\$ is obtained as follows:

$${}^0T_c = {}^0T_{c'} {}^{c'}T_c \quad (5)$$

In particular, the first rigid transformation \${}^0T_{C'}\$ is calculated through geometric measurements taken from a 3D model generated with reverse engineering technique in order to minimize the uncertainty of 3D data. In fact, the adoption of a laser scanner allows to reach accuracy to a tenth of a millimeter. As an alternative to the laser scanner, when a high fidelity between the analytical and the physical prototype is assured,

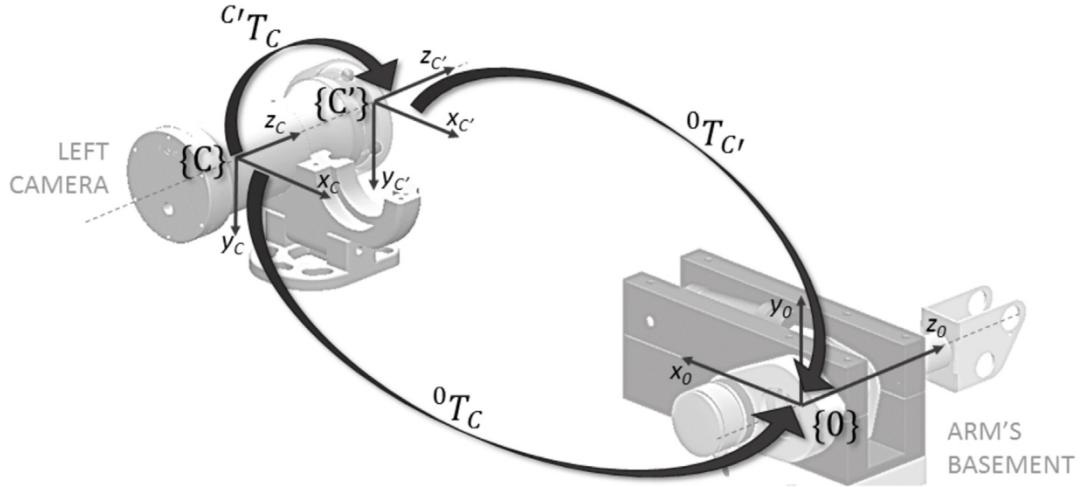


Fig. 9. Geometric approach to determine 0T_c matrix.

the geometric measurements can be taken directly from the CAD model.

About the second transformation matrix ${}^{C'}T_C$, as illustrated in Fig. 10, once the world reference frame $\{W\}$ has been arbitrarily chosen, the rigid transformation from $\{C\}$ to $\{C'\}$ can be defined as follows:

$${}^cT_c = {}^cT_w {}^wT_c = {}^cT_w ({}^cT_w)^{-1} \quad (6)$$

where cT_w matrix is the rigid transformation that brings the word reference frame $\{W\}$ onto the optical-stereo camera reference frame $\{C\}$, and cT_w is the transformation that binds $\{W\}$ to $\{C\}$.

In particular, the cT_w transformation matrix is the exterior orientation of the stereo camera rig computed by means of a calibration procedure that takes into account the geometry of each camera defined according to the standard photogrammetric model. While cT_w matrix is computed by means of an external measurement system consisting of an optical tracking system that allows to reach a sub-mm precision. Nevertheless, to the detriment of more precise and accurate measurements, the position of the optical centre of the camera could be estimated without any specialized equipment, such as a millimeter grid plane (Peer and Solina, 2006). The introduction of the external measurement system entails the splitting of the cT_w transformation matrix in two basic transformations, i.e., ${}^{op}T_w$ and ${}^{op}T_c$, that are calculated with respect to the reference frame of the motion capture camera $\{OP\}$. As a consequence:

$${}^cT_w = ({}^{op}T_{c'})^{-1} {}^{op}T_w = {}^{c'}T_{op} {}^{op}T_w \quad (7)$$

By considering equations (6) and (7), it follows:

$${}^cT_c = {}^cT_w {}^wT_c = {}^{c'}T_{op} {}^{op}T_w {}^wT_c \quad (8)$$

Then finally, the unknown rigid transformation matrix cT_0 is obtained as follows:

$${}^0T_c = {}^0T_{c'} {}^cT_c = {}^0T_{c'} {}^{c'}T_{op} {}^{op}T_w {}^wT_c \quad (9)$$

5.1. Determining the ${}^0T_{C'}$ transformation matrix

The following figure (Fig. 11) shows the resulting 3D point cloud (653'331 points and a density of 4000 points/cm²) of the camera-arm assembly acquired by means of a triangulation based laser scanner (NextEngine 3D Laser Scanner, 2017) placed at a distance of 450 mm from the ROV skid.

Best-fitting algorithms have been applied to the 3D point cloud to fix position and orientation of the local reference frames $\{C'\}$ and $\{0\}$. Once the two reference frames have been defined, the following rigid transformation matrix ${}^0T_{C'}$ has been calculated:

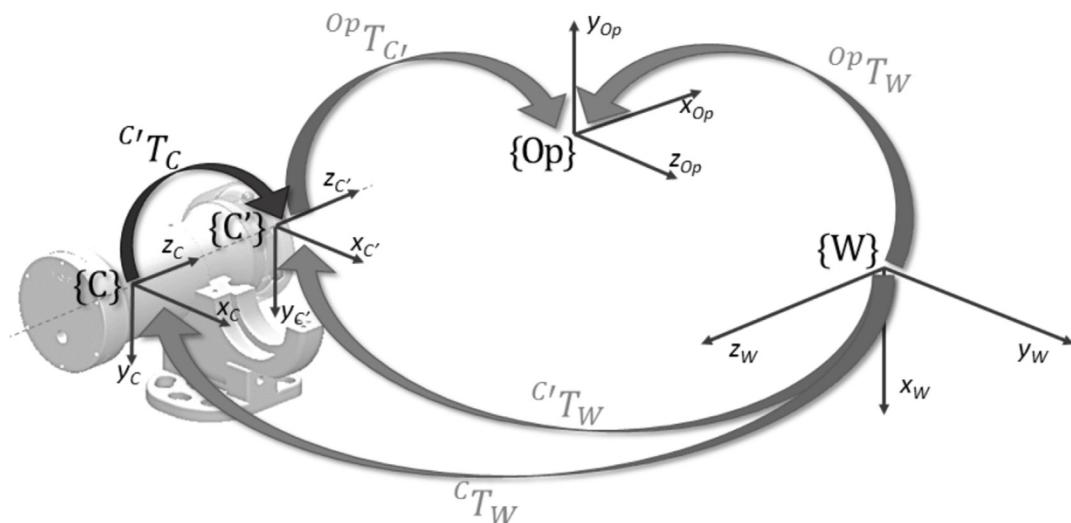


Fig. 10. Geometric approach to determine cT_c matrix.

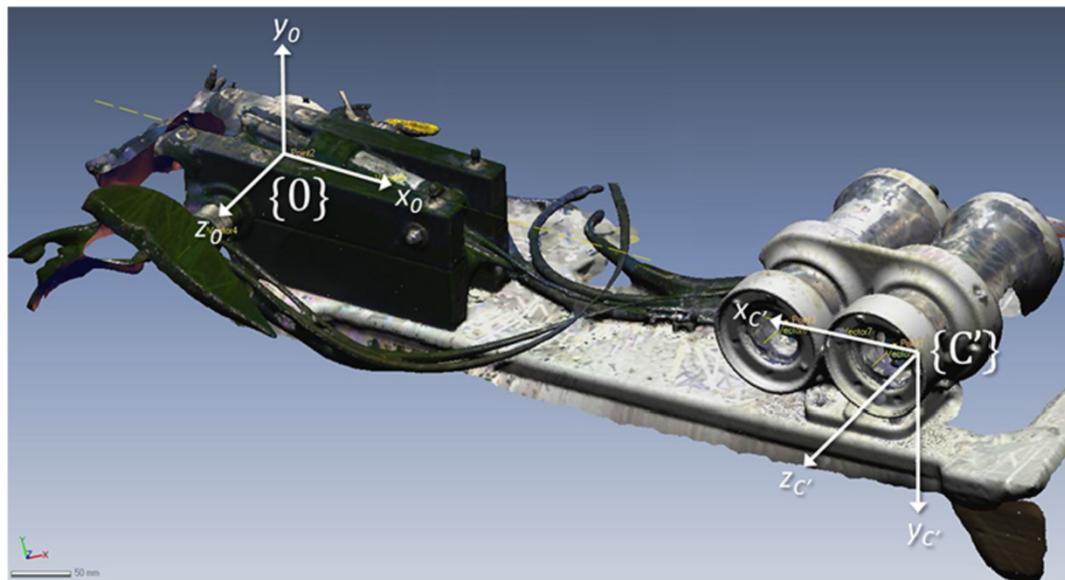


Fig. 11. 3D point cloud of the front side of the ROV skid obtained with laser scanner.

$${}^0T_{c'} = \begin{bmatrix} -0.2249 & -0.9744 & 0 & 679.4500 \\ -0.5359 & 0.1237 & 0.8352 & 38.4700 \\ -0.8138 & 0.1878 & -0.5500 & 125.8000 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

5.2. Determining the ${}^C T_C$ transformation matrix

The set of the internal and external parameters of the optical-stereo camera has been computed by means the calibration method proposed by Zhang (1999), in its version implemented in Matlab by Bouguet (2013). The interior and exterior orientation parameters of each camera have been obtained by correlating the coordinates of known points located on a calibration sample (i.e., a checkerboard) with the corresponding coordinates on the image plane. Subsequently, the exterior orientation of the optical system (stereo calibration) has been computed and the reference frame of the optical-stereo camera $\{C\}$ (whose origin is the perspective centre of the left camera) has been related to the reference frame of the calibration sample $\{W\}$ (whose position and orientation is shown in Fig. 12a). The results of the calibration process are shown in Fig. 12b.

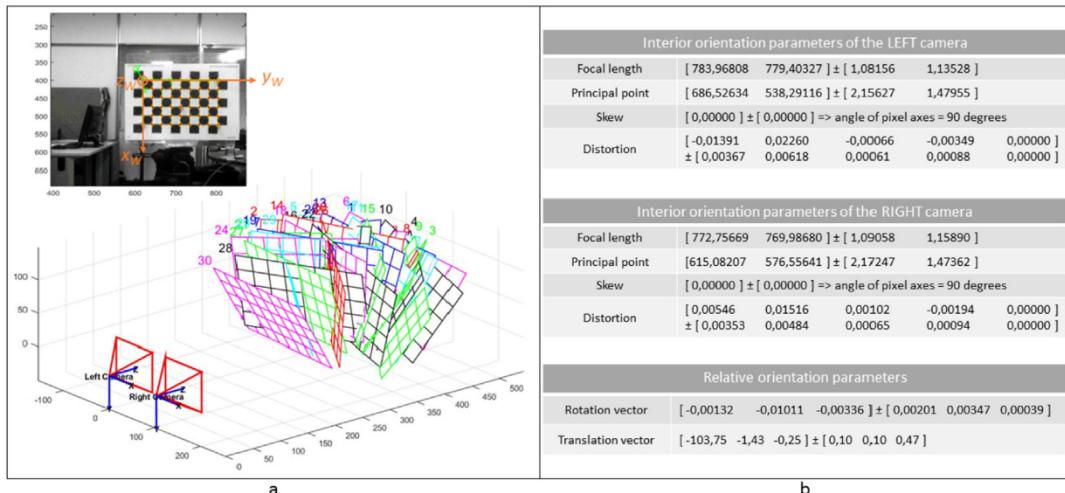


Fig. 12. Results of the calibration process. (a) Pose estimation of the calibration sample with respect to the reference frame of the optical-stereo camera $\{C\}$. (b) Interior and exterior orientation parameters of the stereo rig.

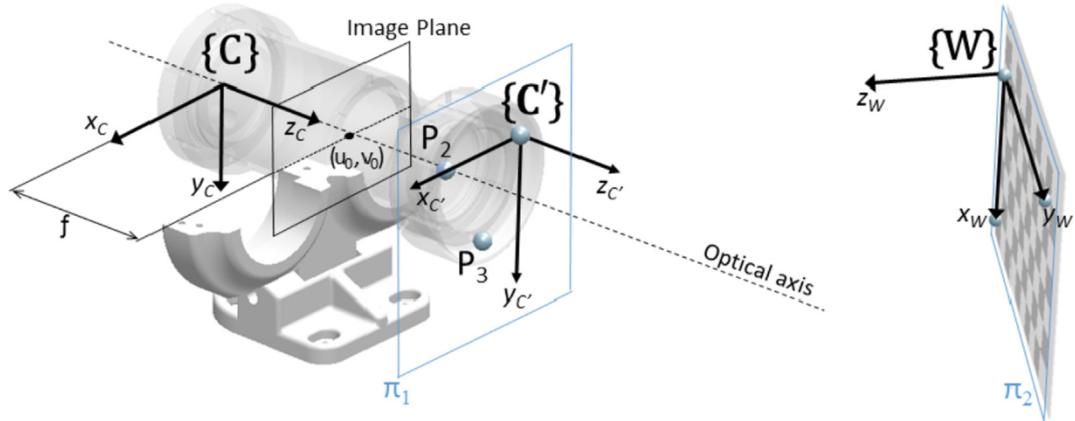


Fig. 13. Schematic layout of the set-up used to calculate the rigid transformation matrix ${}^C T_C$.

sub-mm accuracy level using reflective markers of 7/16 inch (11.112 mm) in diameter.

The reference frame of the camera system $\{C\}$ has been defined with three reflective markers attached to the flange of the flat port of the left camera (Fig. 13). The origin of the coordinate measurement system has been fixed in the point P_1 , with the π_1 plane coincident with the camera flange plane, and the Z-axis along the optical axis. Another set of three reflective markers has been attached to the calibration sample, that consists in a rectangular checkerboard, to define the position and the orientation of the reference frame $\{W\}$. Both the optical-stereo camera (Fig. 14a) and the calibration sample have been mounted on a stable tripod and all infrared (IR) emitting devices and reflective surfaces within the capture volume have been removed or covered prior to perform the measurements (Fig. 14b).

A data record of 0.5 s, that corresponds to 60 3D position values for each marker, has been captured. Then, the 3D coordinates have been fitted through the least squares procedure to find the reference planes of the flange of the camera housing π_1 and of the calibration sample π_2 (Fig. 13).

From the measurements described in the previous steps, the relative pose of the key geometric elements of the camera system is known and the rigid transformation matrix ${}^C T_C$ can be calculated.

Therefore, by adopting equation (8):

$${}^C T_c = \begin{bmatrix} -0.1923 & 0.9800 & 0.010556 & 30.3759 \\ 0.9836 & 0.1895 & -0.2543 & 31.4380 \\ 0.0033 & -0.0526 & 1.0320 & -16.1251 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Once the two transformation matrices, ${}^0 T_{C'}$ and ${}^C T_{C'}$, have been calculated the unknown rigid transformation matrix ${}^C T_0$ can be derived according equation (9):

$${}^C T_c = \begin{bmatrix} -0.9152 & -0.4050 & 0.2355 & 641.9861 \\ 0.2275 & -0.5457 & 0.8011 & 13.2361 \\ 0.3394 & -0.7329 & -0.6600 & 116.8020 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

6. Experimentation

This section presents the experimental results. First, the accuracy of the overall AR visualization system has been estimated through laboratory tests, as described in Section 6.1. Then, the operational capability of the ROV to get closer to a target and to complete a precise cleaning operation by means of an electric brush tool has been assessed in Section 6.2.

6.1. Laboratory tests

The laboratory tests have been carried out at the Department of Mechanical, Energy, and Management Engineering (DIMEG) of the University of Calabria. The purpose of the test has been twofold: verify the capability of the proposed system to generate a depth map with respect to the end-effector position, and to evaluate the accuracy of the overall system.

The experimental set-up adopted in the laboratory tests is depicted in Fig. 15. It consists in the ROV's skid, on which the optical-stereo camera and the robotic arm are mounted, a 3D scanning station, that is a Stonex X300 laser scanner (Stonex, 2017), and various objects of different size, shape, and texture, placed at a distance varying from 0.5 to 3 meters from the skid.

The performance of the overall AR visualization system has been evaluated in three stages. First, the accuracy of the optical-stereo camera in measuring the distance of the objects present within its workspace and in reconstructing their shapes has been assessed. Then,

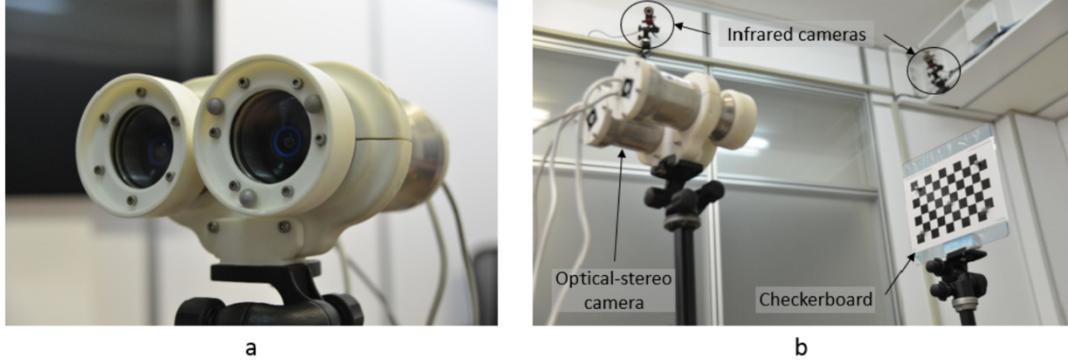


Fig. 14. (a) Optical-stereo camera mounted on tripod with reflective markers, (b) experimental set-up.

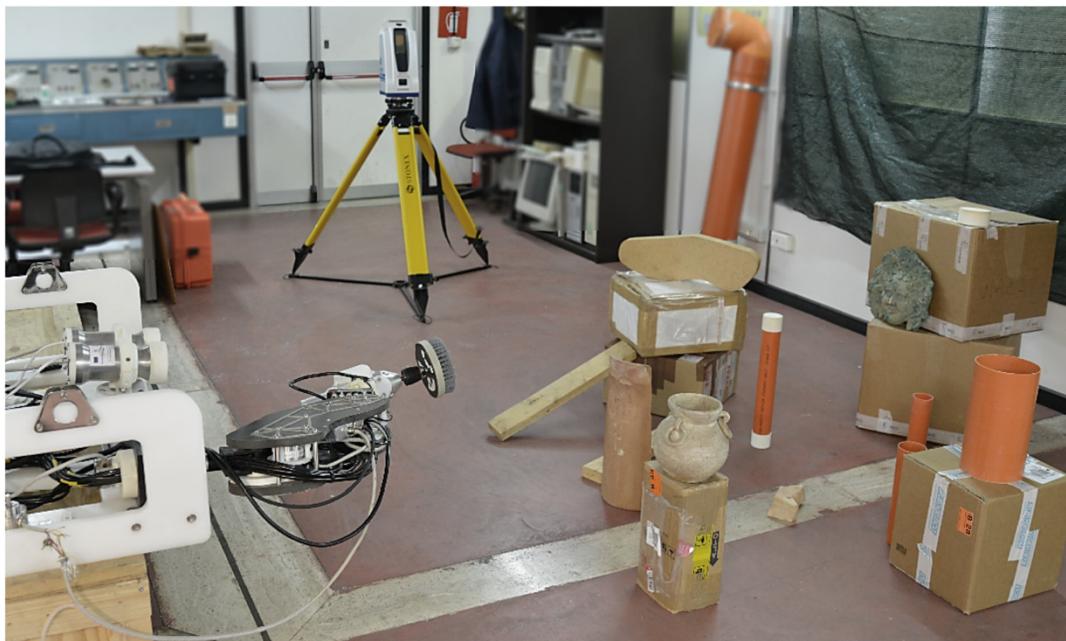


Fig. 15. Experimental set-up.

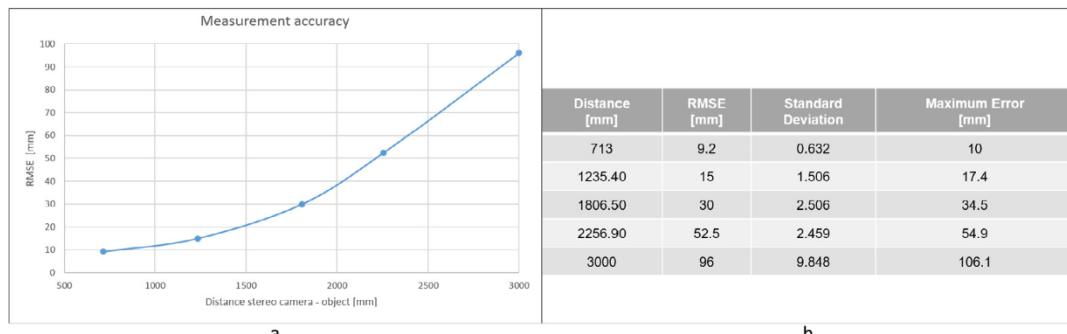


Fig. 16. Results of the optical-stereo camera measurement accuracy. (a) Trend of RMSE values for different distances between the optical-stereo camera and the objects in the workspace. (b) Statistics for accuracy estimation of the optical-stereo camera (mm).

the kinematic performances of the arm has been evaluated. Finally, we have assessed the accuracy obtained on combining the kinematics of the robotic arm and the standard photogrammetric model of the stereo camera to generate a depth map showing the distances of the surface of the scene objects from the end-effector's pose.

6.1.1. Optical-stereo camera accuracy evaluation

The setup of the underwater imaging system with cameras and lens behind a flat port, although easier to manufacturing, can lead to a degradation in accuracy if compared with its performances in air, caused mainly by the refraction of the air-water interface. In particular, as reported in (Menna et al., 2018a,b), with increasing the distance between the entrance pupils of the camera lens and the flat ports of the waterproof cases, the standard photogrammetric model used both in Libelas and OpenCV libraries to perform the on-line AR visualization of the 3D underwater scene represents only a simplification of the underwater image formation process. On the basis of these studies, in order to use the standard photogrammetric model without introducing systematic errors, the cameras have been fixed within the cases through an appropriate support specifically designed to precisely set the distance between the entrance pupil of the camera and the internal surface of the flat port. In particular, a distance of 5 mm, behind a 20 mm thick view port, has been used for this experimentation.

We have experimentally verified in our previous works (Bianco

et al., 2013; Lagudi et al., 2016) that the underwater performances of the optical-stereo camera in the present configuration are comparable with the ones obtained in air (an accuracy degradation ranging from 0.02% to 0.5% has been estimated depending upon the quality of the acquired images) allowing us to assess with a good approximation the operational accuracy of the optical-stereo camera based on the results obtained in the laboratory tests.

Various objects have been acquired at different distances from the optical-stereo camera to assess its measurement accuracy. Since the two calibration methods embedded in the control software (see Section 2.3) achieve equivalent performances in air, the camera has been calibrated using the method proposed by Zhang (see Section 5.2). In particular, 30 point clouds for each pose of the objects have been acquired at a distance between 0.5 m and 3 m from the optical-stereo camera. Then, the distance of the objects has been measured for all the different poses by manually selecting points on the gathered point clouds. For each measurement in the point cloud, the error has been evaluated as the absolute value of the difference between the distance of the object measured by means of the point cloud and those acquired by the laser scanner. The results are depicted in Fig. 16. The Root Mean Square Error (RMSE) grows from values below the centimetre (9.41 mm) when acquiring at short range, to larger values such as 96 mm obtained at longer distances. As expected, due to the short baseline used in the present setup (about 100 mm), the optical-stereo camera give good

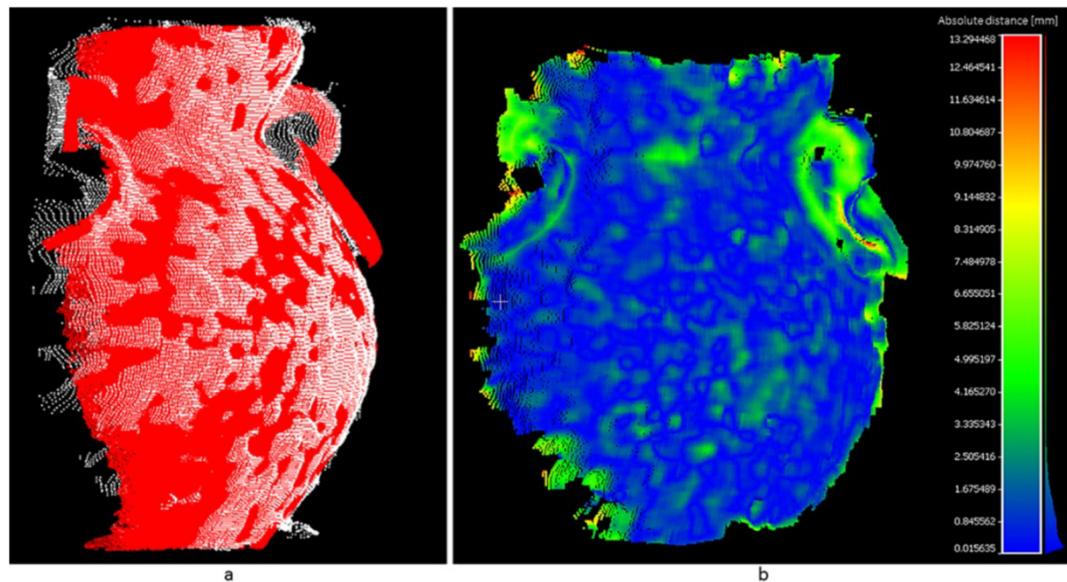


Fig. 17. (a) Reference 3D model overlaid on the gathered point cloud. (b) Gathered point cloud colored according to the distance between each point and reference 3D model.

results in close range acquisition, but deteriorate quickly at longer ranges.

To better assess the performance of the optical-stereo camera, an amphora has been acquired at an approximate distance of 1 m. After manually selecting the points in the cloud that belong to the object, they are registered using Iterative Closest Point (ICP) against reference 3D model obtained using the NextEngine 3D Laser Scanner. The results are presented in Fig. 17. It shows two elements. First, reference 3D model in red overlaid with the gathered point cloud in white (Fig. 17a). Second, the acquired optical cloud where each point is colored according to its distance with respect to reference 3D model (Fig. 17b). The histogram of these distances is shown in Fig. 18. It can be seen that the mean error is around one millimeter (RMSE of 1.57 mm and a standard deviation of 2.18 mm) with a maximum error of 21.20 mm. The resulting error is equivalent to a length accuracy of about 0.5%,

that is consistent with the ones reported in literature (Shortis, 2015).

6.1.2. Kinematic performances of the arm

The forward kinematic model of the robotic arm is affected by errors that could compromise the estimation of the depth map. In particular, the physical properties of the manipulator may differ from the corresponding ideal values due both to geometric and non-geometric errors, such as manufacturing tolerances, assembly errors and misalignment of the joints, elastic deflections of the links, thermal deformations, as well as measurement errors of the encoders itself. These errors can be taken into account by analyzing the kinematic performances of the robotic arm that are expressed in terms of its accuracy (A).

On the basis of the ISO standard 9283:1998 (ISO 9283, 1998), the accuracy (A) can be defined as the ability of the system's end-effector to measure a specified pose. It is measured experimentally by means the

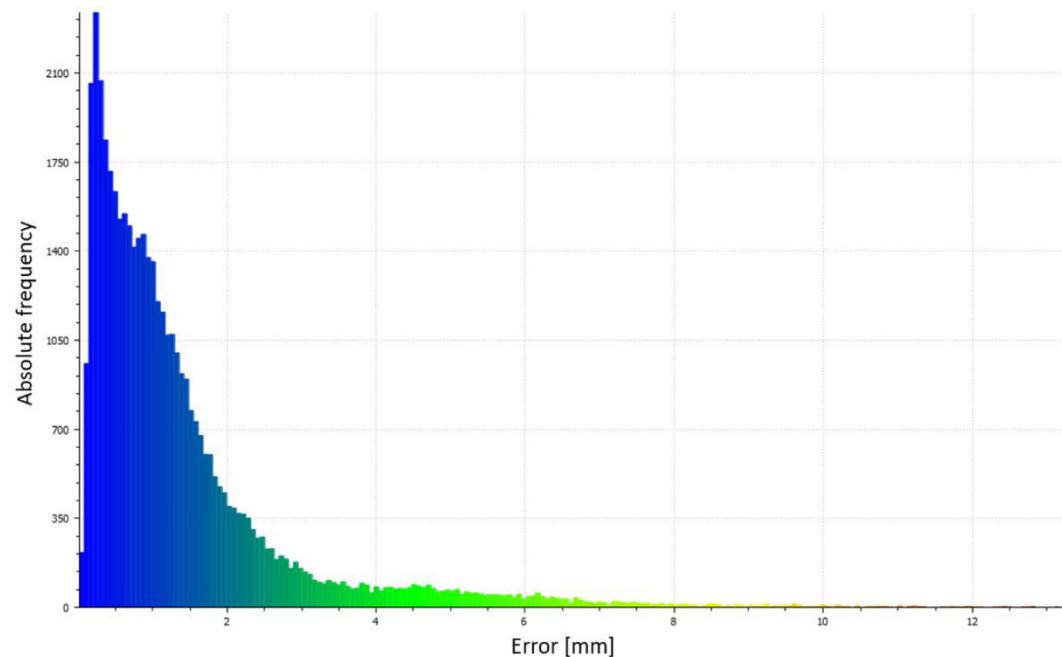


Fig. 18. Error histogram of the Euclidian distance (mm) between the point cloud and reference 3D model representing the acquired amphora.

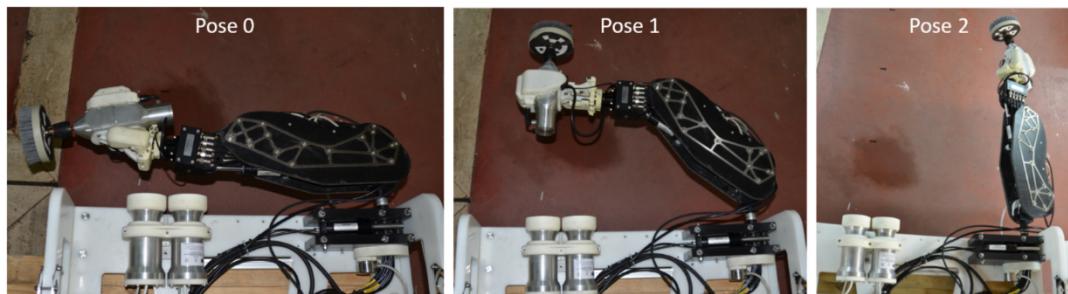


Fig. 19. Three different poses of the robotic arm within its working volume to perform the accuracy estimation.

distribution of the positioning errors calculated by comparing predicted and actual poses of the end-effector arm. In a previous study (Rizzo et al., 2017), we have calculated the error distribution E_A by measuring the end-effector pose in some different configurations, with the resulting estimation of the accuracy of the robotic arm that is $A = \mu_{EA} \pm \sigma_{EA} = 50.62 \pm 24.21$ mm

6.1.3. Accuracy of the AR visualization system

The accuracy of the overall system has been evaluated as the RMSE value between the different distances of the objects in the workspace measured by means of the depth map (as predicted by the mathematical model of the system) and those acquired by the laser scanner (actual distances of the objects with regard to the end-effector' pose). In particular, three series of measurements have been performed, each corresponding to a different pose of the robotic arm (Fig. 19).

After the survey, the acquired 3D point clouds have been processed by means of the software 3D Reconstruction, provided by Stonex, which allows to analyze and elaborate the data in order to calculate the actual distances of the objects from the end-effector pose. The results are depicted in Fig. 20.

As expected, the RMSE values grow when the distance between the optical-stereo camera and the objects in the workspace increases. It is due to the loss of performance of the optical-stereo camera accuracy at longer distances (Fig. 16). The Pose 1 shows higher RMSE values in the workspace. It is due to the stationary deflection caused by gravity that is most severe for this pose of the robotic arm. Anyway, the performance should improve in the underwater environment because of the buoyancy that reduces the gravity effects and allows hydraulic joints to operate more accurately.

The following image (Fig. 21) shows the depth map generated for a specific pose of the end-effector of the robotic arm. To compute the depth map easier to read and interpret, it is visualized as an RGB color gradient map and augmented on the optical flow information provided by the camera. The RGB color scale provides information to interpret the depth map, e.g., the red colors are close objects and blue objects are far away, and its color component range can be customized by specifying the minimum and the maximum distance. The user interface

provides also a set of widgets that allow the operator to interact with the software, through direct manipulation of graphical buttons and menus, in order to edit the main parameters of the stereo camera, customize the transparency of the depth map and manage the layout and number of display windows.

6.2. Field tests

The field tests have been carried out at the headquarters of WASS (Whitehead Alenia Sistemi Subacquei) S.p.A. in Livorno (Italy), one of the partners of the CoMAS project. The tests have been performed in a large pool to gain better control of the environment, where the capability of the ROV to get closer to the target and to complete a precise cleaning operation by means of an electric brush tool was assessed (Fig. 22).

After assembling the Perseo ROV with the skid, on which the manipulator and the optical-stereo camera have been mounted, the vehicle has been immersed in the test pool by means of a gantry crane on rails. The ROV is operated according to a classic procedure in which an operator manages the vehicle and the robotic arm. Typically, the proposed ROV system is controlled by only one pilot. But two persons could possibly be involved for controlling the system, one lead pilot manages vehicle position and a co-pilot operates the underwater manipulator. The depth map provides only visual feedbacks that allows the co-pilot to obtain a reliable estimate of the distance between the end-effector and the target, and then to have an instrument that supports him/her for making decisions while maneuvering the manipulator.

The participants to the test were experienced ROV pilots, involved in the CoMAS project, which provided positive feedbacks and personal opinions about the enhanced situational awareness achieved thanks to the adoption of the proposed AR visualization system.

The following image (Fig. 23) shows the feedbacks provided by the stereo camera while the end-effector is approaching the target.

The software features a high customizability of the user interfaces that allows switching between different visualization modes. In fact, the user has arranged the display windows in order to have 2D and 3D information feedbacks (Fig. 23). In particular, on the right, a display

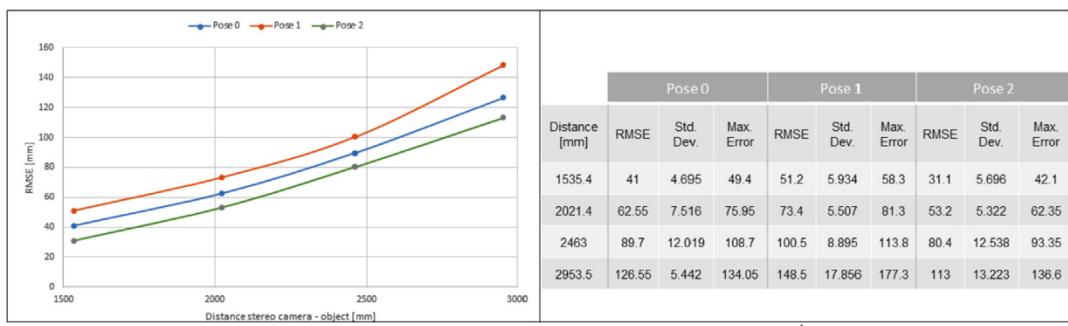


Fig. 20. (a) Comparison of the trend of RMSE values for three poses of the robotic arm when varying the distance between the optical-stereo camera and the objects in the workspace. (b) Statistics for accuracy estimation of the overall system (mm).

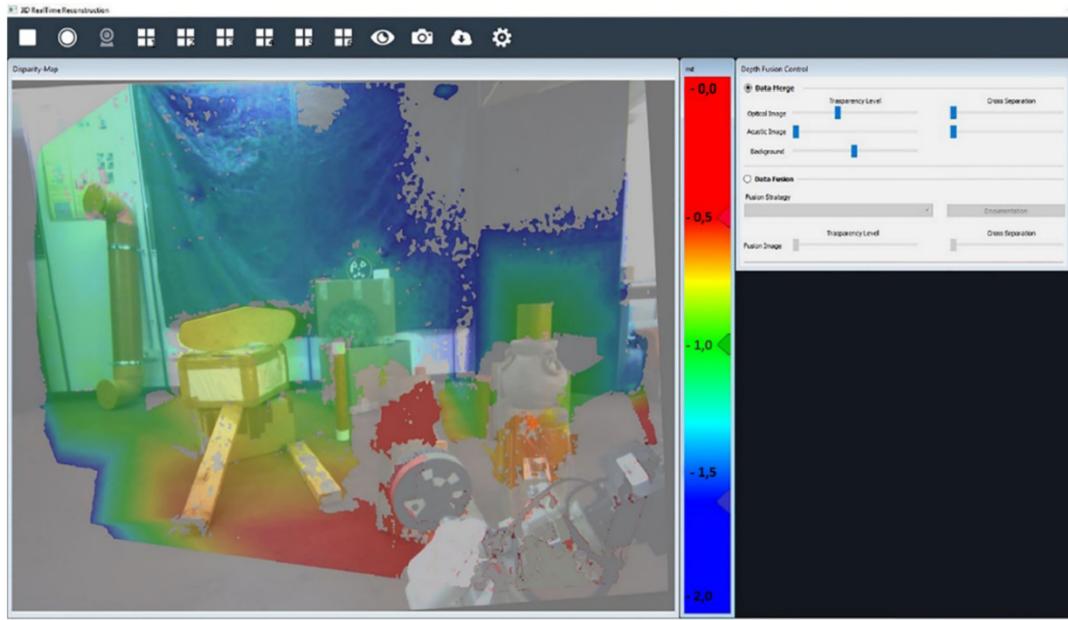


Fig. 21. Depth map augmented on the visual feedback.

window shows the 3D point cloud of the acquired underwater scene, while, on the left, two windows provide to the user respectively the video broadcasted by the left camera and the depth map. Once the vehicle came close to the target, the layout has been rearranged by a user with a single display window (Fig. 24), in which the images broadcasted by the optical-stereo camera are enhanced through the depth map, in order to perform fine manipulation tasks.

The tests have demonstrated that the proposed AR visualization system increases the effectiveness of the control of the robotic underwater arm. In fact, the user can operate safely and accurately thanks to the distance information provided by the optical-stereo camera, instead of piloting the vehicle only by watching the visual feedbacks broadcasted by the camera. The users that have participated in the testing activity made positive remarks about the proposed system for its easily interpretable representation of the visual information and its high flexibility that allows real-time adjustments of camera's parameters in order to cope with the variable conditions of the underwater environment. In this phase of the research project, we were interested to verify

the proper functioning and collect preliminary users' feedback of the proposed AR visualization system. Nevertheless, in future tests, objective metrics would be evaluated with end users in real tasks performed both in freshwater testing pool and in open seawater condition.

7. Conclusions

The paper has presented a novel system based on a sensorized robotic arm, a stereoscopic 3D perception, and augmented reality visualization to provide users a visual feedback of the distance between the end-effector and the underwater scene. The system takes advantage of the modern augmented reality techniques to enrich the information available to the ROV pilots in order to support and improve the overall efficiency of their work. The 3D data of the underwater scene, acquired by the optical-stereo camera mounted on the ROV's skid, are augmented on the visual feedback in order to provide to the pilots a depth map of the distances between the end-effector of the manipulator and the underwater framed scene. As future works, the optical-stereo camera can

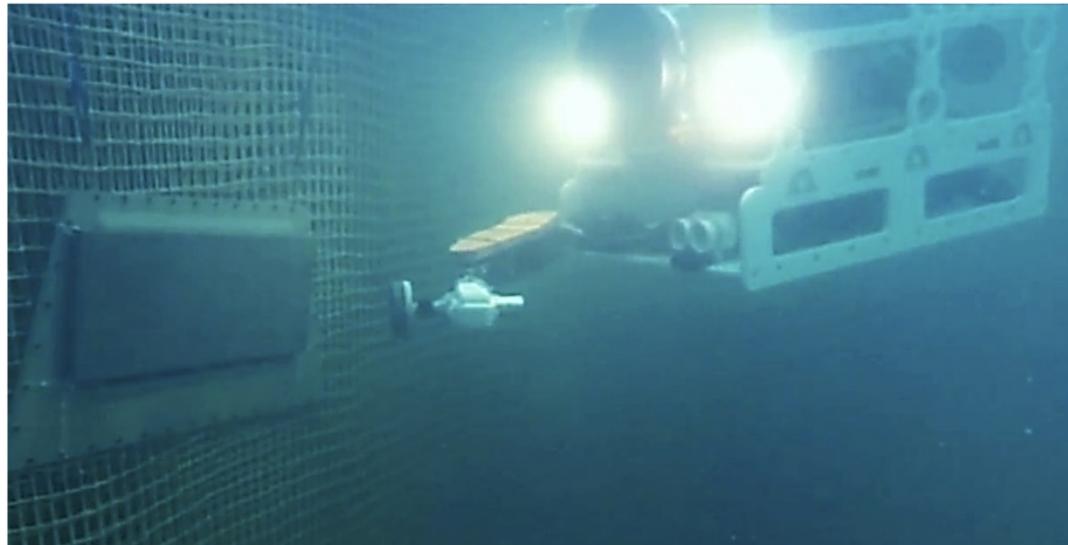


Fig. 22. The end-effector's brush approaches the target.

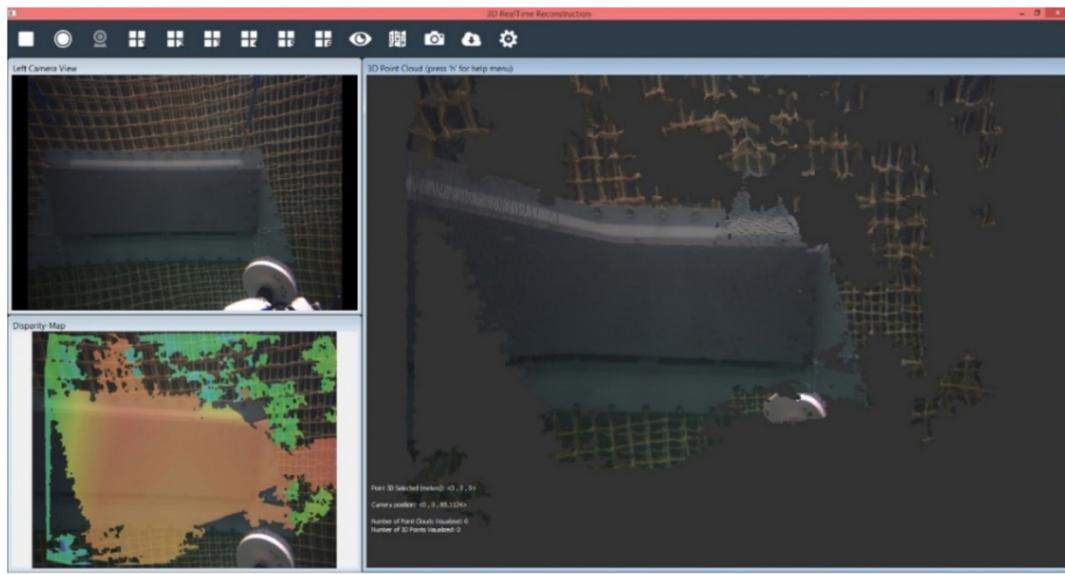


Fig. 23. User interface with different display windows.

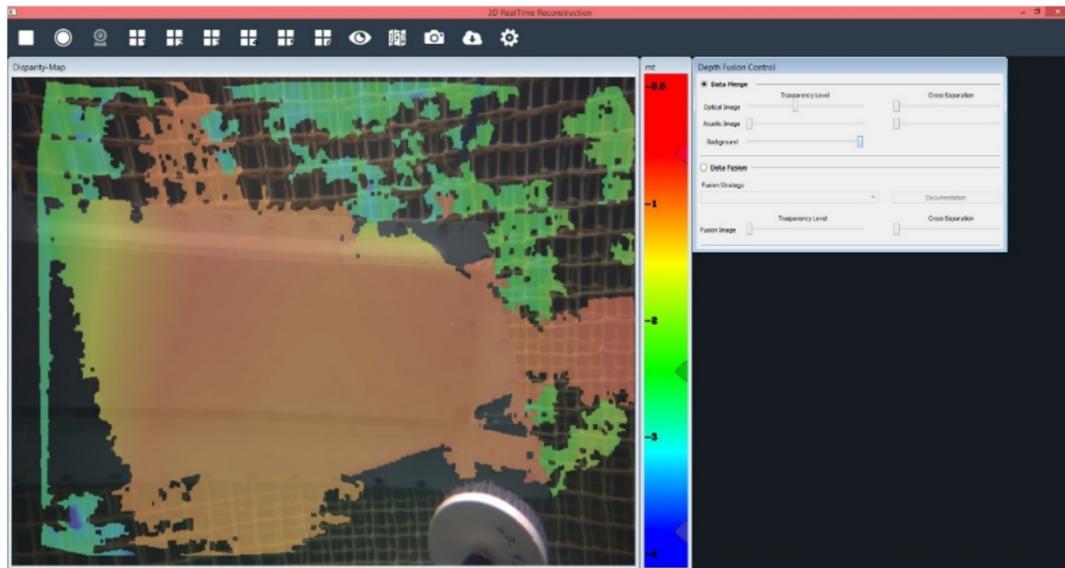


Fig. 24. RGB scalar field window of the camera control software.

be integrated with other sensors capable to calculate the depth map and perform an on-line 3D reconstruction of the underwater scene. For example, could be interesting to evaluate the integration of a 3D acoustic camera (Lagudi et al., 2016) that, unlike the optical sensors, do not suffer from seawater turbidity and allow the pilot to operate also in case of poor visibility conditions.

In considering the limitations and disadvantages derived from the technologies now adopted in the underwater manipulation field, the proposed AR visualization system represents an added value especially in the Underwater Cultural Heritage field where soft touch is required for the manipulation of the archaeological artefacts. In fact, as attested by the ROV pilots that have participated to the first qualitative tests, the proposed system allows a better perception of the position and orientation of the end-effector and its distances from the target objects, and improve the situational awareness of the underwater scene.

Acknowledgements

This work has been supported by the “CoMAS” Project (Ref.: PON01_02140 – CUP: B11C11000600005), financed by the MIUR under the PON ‘R&C’ 2007/2013 (D.D. Prot. n. 01/Ric. 18.1.2010).

References

- Antonelli, G., Caccavale, F., Chiaverini, S., 2004. Adaptive tracking control of underwater vehicle-manipulator systems based on the virtual decomposition approach. *IEEE Trans. Robot. Autom.* 20 (3).
- Barbieri, L., Bruno, F., De Napoli, L., Gallo, A., Muzzupappa, M., 2017. Design of an Electric Tool for Underwater Archaeological Restoration Based on a User Centred Approach. *Advances on Mechanics, Design Engineering and Manufacturing*. Springer International Publishing, pp. 353–362.
- Bianco, G., Gallo, A., Bruno, F., Muzzupappa, M., 2013. A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. *Sensors* 13 (8), 11007–11031.
- Bouguet, 2013. Camera calibration toolbox for Matlab. 01.08.18. http://www.vision.caltech.edu/bouguetj/calib_doc/.
- Brown, D.C., 1971. Close-range camera calibration. *Photogramm. Eng.* 37 (8), 855–866.
- Bruno, F., Gallo, A., Barbieri, L., Muzzupappa, M., Ritacco, G., Lagudi, A., La Russa, M.F.,

- Ruffolo, S.A., Crisci, G.M., Ricca, M., Comite, V., Davide, B., Stefano, Di, Guida, R., 2016. The CoMAS project: new materials and tools for improving the in-situ documentation, restoration and conservation of underwater archaeological remains. *Mar. Technol. Soc. J.* 50 (4), 108–118.
- Bruno, F., Muzzupappa, M., Gallo, A., Barbieri, L., Spadafora, F., Galati, D., Davidde, B., Petriaggi, R., 2015a. Electromechanical devices for supporting the restoration of underwater archaeological artefacts. In: International Conference. OCEANS, Genova.
- Bruno, F., Muzzupappa, M., Lagudi, A., Gallo, A., Spadafora, F., Ritacco, G., Angilica, A., Barbieri, L., Di Stefano, G., Guida, R., Di Lecce, N., Saviozzi, G., Laschi, C., 2015b. A ROV for supporting the planned maintenance in underwater archaeological sites. In: Proceeding of IEEE/MTS OCEANS Conference. Genova (Italy).
- Bruno, F., Bianco, G., Muzzupappa, M., Barone, S., Razionale, A.V., 2011. Experimentation of structured light and stereo vision for underwater 3D reconstruction. *Journal of Photogrammetry and Remote Sensing, ISPRS* 66 (4), 508–518.
- CoMAS project, 2013. In-situ conservation planning of underwater archaeological artefacts. 01.08.18. <http://www.comasproject.eu>.
- Denavit, J., Hartenberg, R.S., 1955. A kinematic notation for lower-pair mechanisms based on matrices. *J. Appl. Mech.* 77 (2), 215–221.
- Galloway, K.C., Becker, K.P., Phillips, B., Kirby, J., Licht, S., Tchernov, D., Wood, R.J., Gruber, D.F., 2016. Soft robotic grippers for biological sampling on deep reefs. *Soft Robot.* 3 (1), 23–33.
- Geiger, A., Roser, M., Urtasun, R., 2011. Efficient large-scale stereo matching. In: Computer Vision-ACCV 2010. Springer, Berlin Heidelberg, pp. 25–38.
- ISO 9283, 1998. Manipulating Industrial Robots - Performance Criteria and Related Test Methods. Standard. International Organization for Standardization, Geneva, Switzerland.
- Lagudi, A., Bianco, G., Muzzupappa, M., Bruno, F., 2016. An alignment method for the integration of underwater 3D data captured by a stereovision system and an acoustic camera. *Sensors* 16 (4), 536.
- Marani, G.C., Yuh, J., 2009. Underwater autonomous manipulation for intervention missions AUVs. *Ocean Eng.* (36), 15–23.
- Menna, F., Nocerino, E., Remondino, F., 2018a. Photogrammetric modelling of submerged structures: influence of underwater environment and lens ports on three-dimensional (3D) measurements. In: Remondino, F., Georgopoulos, A., González-Aguilar, D., Agrafiotis, P. (Eds.), *Latest Developments in Reality-based 3d Surveying and Modelling*. MDPI, Basel, Switzerland, pp. 279–303 2018.
- Menna, F., Nocerino, E., Drap, P., Remondino, F., Murtiyoso, A., Grussenmeyer, P., Börlin, N. 2018. Improving underwater accuracy by empirical weighting of image observations. In: ISPRS Technical Commission II Symposium 2018, Riva del Garda, Italy, June 3–7, 2018 (vol. 42), pp. 699–705.
- Mohan, S., Kim, J., 2015. Coordinated motion control in task space of an autonomous underwater vehicle-manipulator system. *Ocean Eng.* 104, 155–167.
- NaturalPoint Inc, 2017. OptiTrack Flex 13. 01.08.18. <http://www.naturalpoint.com/optitrack/products/tracking-tools-bundles>.
- NextEngine 3D Laser Scanner, 2017. <<http://www.nextengine.com>> (accessed 01.08.18).
- Peer, P., Solina, F., 2006. Where physically is the optical center? *Pattern Recogn. Lett.* 27 (10), 1117–1121.
- Prats, M., Fernandez, J., Sanz, P., 2012. An approach for semi-autonomous recovery of unknown objects in underwater environments. In: IEEE Int. Conf. Optimization of Electrical and Electronic Equipment.
- Rahman, T., Krouglicof, N., 2012. An efficient camera calibration technique offering robustness and accuracy over a wide range of lens distortion. *IEEE Image process* 21, 626–637.
- Rizzo, D., Bruno, F., Barbieri, L., Muzzupappa, M., 2017. Kinematic performances evaluation of a hydraulic underwater manipulator. In: Proceeding of IEEE/MTS OCEANS Conference. Aberdeen (UK).
- Sagara, S., Ambar, R., 2015. Development of a master controller for a 3 link dual arm underwater robot. *Artif. Life Robot.* 20, 327–335.
- Sakagami, N., Shibata, M., Inoue, T., 2010. Development of a Human-sized ROV with Dual-arm. *Oceans Sydney*.
- Shim, H., Jun, B.-H., Lee, P.-M., Baek, H., Lee, J., 2010. Workspace control system of underwater tele-operated manipulators on a ROV. *Ocean Eng.* 37, 1036–1047.
- Shortis, M., 2015. Calibration techniques for accurate measurements by underwater camera systems. *Sensors* 15 (12), 30810–30826.
- Skarlatos, D., Agapiou, A., Rova, M., 2010. Photogrammetric support on an underwater archaeological excavation site: the Mazotos shipwreck case. *Euromed 2010*, Limassol, Cyprus 8–11.
- Stonex, 2017. X300 laser scanner. 01.08.18. <http://www.stonexpositioning.com/index.php/en/product/laser-scanners/x300-detail>.
- Zhang, J., Li, W., Yu, J., Feng, X., Zhang, Q., Chen, G., 2017. Study of manipulator operations maneuvered by a ROV in virtual environments. *Ocean Eng.* 142, 292–302.
- Zhang, Z., 1999. Flexible camera calibration by viewing a plane from unknown orientations. In: Computer Vision, the Proceedings of the Seventh IEEE International Conference on, vol. 1. pp. 666–673.